The List-Decoding Size of Reed-Muller Codes

Tali Kaufman *
MIT
kaufmant@mit.edu

Shachar Lovett †
Weizmann Institute of Science
shachar.lovett@weizmann.ac.il

November 14, 2008

Abstract

In this work we study the list-decoding size of Reed-Muller codes. Given a received word and a distance parameter, we are interested in bounding the size of the list of Reed-Muller codewords that are within that distance from the received word. Previous bounds of Gopalan, Klivans and Zuckerman [4] on the list size of Reed-Muller codes apply only up to the minimum distance of the code. In this work we provide asymptotic bounds for the list-decoding size of Reed-Muller codes that apply for *all* distances. Additionally, we study the weight distribution of Reed-Muller codes. Prior results of Kasami and Tokura [8] on the structure of Reed-Muller codewords up to twice the minimum distance, imply bounds on the weight distribution of the code that apply only until twice the minimum distance. We provide accumulative bounds for the weight distribution of Reed-Muller codes that apply to *all* distances.

1 Introduction

The problem of list-decoding an error correcting code is the following: given a received word and a distance parameter find all codewords of the code that are within the given distance from the received word. List-decoding is a generalization of the more common notion of unique decoding in which the given distance parameter ensures that there can be at most one codeword of the code that is within the given distance from the received word. The notion of list-decoding has numerous practical and theoretical implications. The breakthrough results in this field are due to Goldreich and Levin [3] and Sudan [10] who gave efficient list decoding algorithms for the Hadamard code and the Reed-Solomon code. See surveys by Guruswami [5] and Sudan [11] for further details. In complexity, list-decodable codes are used to perform hardness amplification of functions [12]. In cryptography, list-decodable codes are used to construct hard-core predicates from one way functions [3]. In learning theory, list decoding of Hadamard codes implies learning parities with noise [7].

In this paper we study the question of list-decoding Reed-Muller codes. Specifically, we are interested in bounding the list sizes obtained for different distance parameters for the list-decoding problem.

*Research supported in part by NSF Awards CCF-0514167 and NSF-0729011.

†Research supported partly by the Israel Science Foundation (grant 1300/05). Research was conducted partly when the author was an intern at Microsoft Research.

Reed-Muller codes are very fundamental and well studied codes. $RM(n,d)$ is a linear code, whose codewords $f \in RM(n,d) : \mathbb{F}_2^n \to \mathbb{F}_2$ are evaluations of polynomials in n variables of total degree at most d over $\mathbb{F}_2$. In this work we study the code $RM(n,d)$ when $d \ll n$, and are interested in particular in the case of constant d.

The following facts regarding $RM(n,d)$ are straight-forward: It has block length of 2^n, dimension $\sum_{i \leq d} \binom{n}{i}$ and minimum relative distance $\frac{2^{n-d}}{2^n} = 2^{-d}$. We define:

Definition 1 (Relative weight of a function)**.** The relative weight of a function/codeword $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is the fraction of non-zero elements,

$$wt(f) = \frac{1}{2^n} |\{x \in \mathbb{F}_2^n : f(x) = 1\}|$$

A closely related definition is the distance between two functions

Definition 2 (Relative distance between two functions)**.** The relative distance between two functions $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$ is defined as

$$dist(f,g) = \mathbb{P}_{x \in \mathbb{F}_2^n}[f(x) \neq g(x)]$$

The main focus of this work is in understanding the asymptotic growth of the list size in list-decoding of Reed-Muller codes, as a function of the distance parameter. Specifically we are interested in obtaining bounds on the following.

Definition 3 (List-decoding size)**.** For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ let the ball at relative distance α around f be

$$B(f,\alpha) = \{p \in RM(n,d) : dist(p,f) \leq \alpha\}$$

The list-decoding size of $RM(n,d)$ at distance α, denoted by $L(\alpha)$, is the maximal size of $B(f,\alpha)$ over all possible functions f, i.e.

$$L(\alpha) = \max_{f:\mathbb{F}_2^n \to \mathbb{F}_2} |B(f,\alpha)|$$

In a recent work Gopalan, Klivans and Zuckerman [4] prove that for distances up to the minimal distance of the code, the list-decoding size of Reed-Muller codes remains constant.

Theorem 1 (Theorem 11 in [4])**.**

$$L(2^{-d} - \epsilon) \leq O\left((1/\epsilon)^{8d}\right)$$

Their result of bounding the list-decoding size of Reed-Muller codes is inherently limited to work up to the minimum distance of the code, since it uses a structural theorem of Kasami and Takura on Reed-Muller codes [8], which implies a bound on the weight distribution of Reed-Muller codes that works up to twice the minimum distance of the code.

Additionally, the work of [4] has developed a list-decoding algorithm for $RM(n,d)$ whose running time is polynomial in the worst list-decoding size and in the block length of the code.

Theorem 2 (Theorem 4 in [4])**.** *Given a distance parameter α and a received word $R : \mathbb{F}_2^n \to \mathbb{F}_2$, there is an algorithm that runs in time poly$(2^n, L(\alpha))$ and produces a list of all $p \in RM(n,d)$ such that $dist(p,R) \leq \alpha$.*

Since Gopalan et al. could obtain non-trivial bounds on the list-decoding size for distance parameter α that is bounded by the minimum distance of the Reed-Muller code, their algorithm yields meaningful running time only for α that is less than twice the minimum distance of the code.

1.1 Weight distribution of Reed-Muller codes

A close notion to the list-decoding size of Reed-Muller code is the weight distribution of the code.

Definition 4 (Accumulative weight distribution)**.** The accumulative weight distribution of $RM(n,d)$ at a relative weight α is the number of codewords up to this weight, i.e.

$$A(\alpha) = |\{p \in RM(n,d) : wt(p) \leq \alpha\}|$$

where $0 \leq \alpha \leq 1$.

It is well-known that for any $p \in RM(n,d)$ which is not identically zero, $wt(p) \geq 2^{-d}$. Thus, $A(2^{-d} - \epsilon) = 1$ for any $\epsilon > 0$. Kasami and Tokura [8] characterized the codewords in $RM(n,d)$ of weight up to twice the minimal distance of the code (i.e up to distance 2^{1-d}). Based on their characterization one could conclude the following.

Corollary 3 (Corollary 10 in [4])**.**

$$A(2^{1-d} - \epsilon) \leq (1/\epsilon)^{2(n+1)}$$

Corollary 3 and simple lower bounds (which we show later, see Lemma 8) show that $A(\alpha) = 2^{\Theta(n)}$ for $\alpha \in [2^{-d}, 2^{1-d} - \epsilon]$ for any $\epsilon > 0$ (and constant d).

1.2 Our Results

Gopalan et al. [4] left as an open problem the question of bounding the list-decoding size of Reed-Muller codes beyond the minimal distance. In particular, they ask what is the maximal α s.t. $L(\alpha) = 2^{O(n)}$.

In this work we answer their question. Specifically we show bounds on the list-decoding size of Reed-Muller code for distances passing the minimal distance. In fact, we show that the asymptotic behavior of $L(\alpha)$, for all $0 \leq \alpha \leq 1$. Our first result shows that there exist "cut-off distances", at which the list-decoding size changes from $2^{\Theta(n^\ell)}$ to $2^{\Theta(n^{\ell+1})}$:

Theorem 4 (First main theorem - list-decoding size)**.** *Let $1 \leq \ell \leq d-1$ be an integer, and let $\epsilon > 0$. For any $\alpha \in [2^{\ell-d-1}, 2^{\ell-d} - \epsilon]$*

$$L(\alpha) = 2^{\Theta(n^\ell)}$$

and $L(\alpha) = 2^{\Theta(n^d)}$ for any $\alpha \geq 1/2$.

Using Theorem 4, and Theorem 2 we obtain the following algorithmic result for list-decoding Reed-Muller codes from an arbitrary distance.

Theorem 5 (List-decoding algorithm)**.** *Given a received word $R : \mathbb{F}_2^n \to \mathbb{F}_2$ that is at distance α from $RM(n,d)$, for $\alpha \in [2^{\ell-d-1}, 2^{\ell-d} - \epsilon]$. where $1 \leq \ell \leq d-1$ is an integer, and $\epsilon > 0$. There exists an algorithm that runs in time poly$(2^{\Theta(n^\ell)})$ and produces a list of all $p \in RM(n,d)$ such that $dist(p,R) \leq \alpha$*

The weight distribution of $RM(n,d)$ codes beyond twice the minimum distance was widely open prior to our work. See e.g. Research Problem (15.1) in [9] and the related discussion in that Chapter.

In this work we provide asymptotic bounds for the weight distribution of $RM(n,d)$ that applied for all weights $2^{-d} \leq \alpha \leq 1/2$. Specifically, our second main result gives exact boundaries on the range of α for which $A(\alpha) = 2^{\Theta(n^\ell)}$, for any $\ell = 1, 2, ..., d$.

Theorem 6 (Second main Theorem - accumulative weight distribution)**.** *Let $1 \leq \ell \leq d-1$ be an integer, and let $\epsilon > 0$. For any $\alpha \in [2^{\ell-d-1}, 2^{\ell-d} - \epsilon]$*

$$A(\alpha) = 2^{\Theta(n^\ell)}$$

and $A(\alpha) = 2^{\Theta(n^d)}$ for any $\alpha \geq 1/2$.

Theorems 4 and 6 are asymptotically tight for constant $\epsilon > 0$. For sub-constant ϵ, and $\alpha \in [2^{\ell-d-1}, 2^{\ell-d} - \epsilon]$, our bound gives:

$$A(\alpha) \leq L(\alpha) \leq 2^{O(n^\ell/\epsilon^2)}$$

We conjecture this dependency on ϵ is not optimal, and the correct dependency should be $\log(1/\epsilon)$ instead of $1/\epsilon^2$. We expand more on that in the body of the paper.

1.3 Techniques

The bounds on the accumulative weight distribution of the Reed-Muller code are obtained using the following novel strategy. We show that a function $f : F_2^n \to F_2$ whose weight is bounded by $wt(f) \leq 2^{-k}(1-\epsilon)$ can be *computed* as an expectation of its kth-derivatives multiplied by some bounded coefficients (Lemma 10).

Using standard sampling methods we then show (Lemma 11) that a function $f : F_2^n \to F_2$ whose weight is bounded by $wt(f) \leq 2^{-k}(1-\epsilon)$ can be well approximated by a constant number $c = c(k, \epsilon)$ of its kth-derivatives. This implies that every $RM(n, d)$ codeword of weight up to $2^{-k}(1-\epsilon)$ can be well approximated by $c = c(k, \epsilon)$ of its kth-derivatives. Since the distance between every pair of $RM(n, d)$ codewords is at least 2^{-d}, a good enough approximation of a $RM(n, d)$ codeword determines the Reed-Muller codeword uniquely. Hence, the number of $RM(n, d)$ codewords up to weight $2^{-k}(1-\epsilon)$, is bounded by the number of kth-derivatives to the power of $c = c(k, \epsilon)$. As $RM(n, d)$ codewords are polynomials of degree at most d, their kth-derivatives are polynomials of degree at most $d-k$. There can be at most $\Theta(2^{n^{d-k}})$ such derivatives. Thus, the number of $RM(n, d)$ codewords up to weight $2^{-k}(1-\epsilon)$, can be bounded by $O(2^{n^{d-k}})^c = O(2^{c \cdot n^{d-k}})$. We complement these upper bound estimations with matching lower bounds.

A similar work in this line is the work of Viola and Bogdanov [2], which shows that a function $f : F_2^n \to F_2$ whose weight is bounded by $wt(f) \leq 1/2 - \epsilon$ can be well approximated by $c = c(k, \epsilon)$ of its 1st-derivatives. Note that approximation by 1st-derivatives *does not* imply in general approximation by kth-derivatives which is crucial for obtaining our bounds here.

The bounds on the list-decoding size of Reed-Muller codes are obtained using similar techniques to the ones used for bounding the accumulative weight distributions.

1.4 Generalized Reed-Muller Codes

The problems of bounding both the accumulative weight distribution and the list-decoding size can be extended to Generalized Reed-Muller, the code of low-degree polynomials over larger fields. However, our techniques fail to prove tight result in these cases. We provide some partial results for this case and make a conjecture about the correct bounds in Appendix A.

1.5 Organization

Although our goal is bounding the list-decoding size of Reed-Muller codes, we first study the accumulative weight distribution of Reed-Muller codes. The techniques we develop are then easily transferred to bounding also the list-decoding size.

The paper is organized as follows. In Section 2 we study the weight distribution of Reed-Muller codes and we prove the Second Main Theorem (Theorem 6). In Section 3 study the list-decoding size of Reed-Muller codes. We generalize the techniques of Section 2 to prove the First Main Theorem (Theorem 4). In Section A we study similar questions for Generalized Reed-Muller code and provide non-tight bounds for these codes.

2 Weight distribution of Reed-Muller codes

In this section we study the weight distribution of Reed-Muller codes, and we prove our Second Main Theorem (Theorem 6). Let $RM(n,d)$ stand for the code of multivariate polynomials $p(x_1,...,x_n)$ over $\mathbb{F}_2$ of total degree at most d. In the following n and d will always stand for the number of variables and the total degree. We will assume that $d \ll n$, and study in particular the case of constant d.

Our Second Main Theorem (Theorem 6) is a direct corollary of Theorem 7, giving an upper bound on the accumulative weight at distance $2^{\ell-d}-\epsilon$, and Lemma 8, giving a simple lower bound at distance $2^{\ell-d-1}$.

Theorem 7 (Upper bound on the accumulative weight)**.** *For any integer $1 \le k \le d-1$,*

$$A(2^{-k}(1-\epsilon)) \le c_1 2^{c_2 \frac{n^{d-k}}{\epsilon^2}}$$

where $c_1 = (1/\epsilon)^{O(d/\epsilon^2)}$ and $c_2 = O(d/(d-k)!)$. Importantly, c_1, c_2 are independent of n, and c_2 is independent of ϵ. In particular for constant d we get that

$$A(2^{-k}-\epsilon) \le 2^{O(\frac{n^{d-k}}{\epsilon^2})}$$

Lemma 8 (Lower bound on the accumulative weight)**.** *For any integer $1 \le k \le d$*

$$A(2^{-k}) \ge 2^{\frac{n^{d-k+1}}{(d-k+1)!}(1+o(1))}$$

In the upper bound on $A(\alpha)$, while the dependence on n is tight, we believe the dependence on ϵ can be improved. For $k = d-1$ (and constant d), the characterization of [8] shows that

$$A(2^{1-d}-\epsilon) = 2^{\Theta(n\log(1/\epsilon))}$$

We conjecture that this is the correct dependence on ϵ in all the range:

Conjecture 9. *Let d be constant. For any integer $1 \le k \le d-1$,*

$$A(2^{-k}-\epsilon) = 2^{\Theta(n^{d-k}\log(1/\epsilon))}$$

We start by proving the lower bound.

Proof of Lemma 8. Single out k variables $x_1, ..., x_k$, and let q be any degree $d-k+1$ polynomials on the remaining $n-k$ variables. First, for any such q, the following degree d polynomial has relative weight exactly 2^{-k}:

$$q'(x_1, ..., x_n) = x_1 x_2 ... x_{k-1}(x_k + q(x_{k+1}, ..., x_n))$$

The number of different polynomials q is

$$2^{\binom{n-k}{d-k+1}} = 2^{\frac{n^{d-k+1}}{(d-k+1)!}(1+o(1))}$$

□

We will prove Theorem 7 in the rest of the section. We start by defining discrete derivatives, which will be our main tool in the proof.

Definition 5. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ by a function. We define the discrete derivative of f in direction $a \in \mathbb{F}_2^n$ to be

$$f_a(x) = f(x+a) + f(x)$$

We define the iterated discrete derivative of f in directions $a_1, ..., a_k \in \mathbb{F}_2^n$ to be

$$f_{a_1,...,a_k}(x) = (...((f_{a_1})_{a_2})...)_{a_k}(x) = \sum_{S \subseteq [k]} f(x + \sum_{i \in S} a_i)$$

We note that usually derivatives are defined as $f_a(x) = f(x+a) - f(x)$, but since we are working over $\mathbb{F}_2$, we can ignore the signs.

We define another notion which is central to our proof, namely the bias of a function.

Definition 6. The bias of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is

$$bias(f) = \mathbb{E}_{x \in \mathbb{F}_2^n}[(-1)^{f(x)}] = \mathbb{P}[f=0] - \mathbb{P}[f=1] = 1 - 2wt(f)$$

The following lemma will be the heart of our proof. It shows that if a function f has weight less than 2^{-k}, then it can be computed by a its iterated k-derivatives.

Lemma 10 (Main technical lemma)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function s.t. $wt(f) < 2^{-k}(1-\epsilon)$. Then the function $(-1)^{f(x)} : \mathbb{F}_2^n \to \{-1, 1\}$ can be written as*

$$(-1)^{f(x)} = \mathbb{E}_{a_1,...,a_k \in \mathbb{F}_2^n}[\alpha_{a_1,...,a_k}(-1)^{f_{a_1,...,a_k}(x)}]$$

where $\alpha_{a_1,...,a_k}$ are real numbers, of absolute value of at most $\frac{10}{\epsilon}$

We will first prove Theorem 7 given Lemma 10, and then turn to prove Lemma 10. We will also need the following well-known technical lemma, which shows how to transform calculation by averaging many functions, to approximation by averaging few functions.

Lemma 11 (Approximation by sampling)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function, $H = \{h_1, ..., h_t\}$ a set of functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2$, s.t. there exist constants $c_{h_1}, ..., c_{h_t}$ of absolute value at most C, s.t.*

$$(-1)^{f(x)} = \mathbb{E}_{i \in [t]}[c_{h_i}(-1)^{h_i(x)}] \qquad (\forall x \in \mathbb{F}_2^n)$$

Then f can be approximated by a small number of the functions $h_1, ..., h_t$. For any $\delta > 0$, there exist functions $h_1, ..., h_\ell \in H$ for $\ell = O(C^2 \log 1/\delta)$, and a function $F : \mathbb{F}_2^\ell \to \mathbb{F}_2$, s.t. the relative distance between $f(x)$ and $F(h_1(x), ..., h_\ell(x))$ is at most δ, i.e.

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[f(x) \neq F(h_1(x), ..., h_\ell(x))] \leq \delta$$

The function F is a weighted majority, i.e. it is of the form:

$$F(h_1(x), ..., h_\ell(x)) = sign(\frac{\sum_{i=1}^{\ell} s_i(-1)^{h_i(x)}}{\ell})$$

where $sign(x)$ is defined by $sign(x) = 1$ if $x \geq 0$ and $sign(x) = -1$ if $x < 0$. Moreover, we can have $s_1, ..., s_\ell$ to be integers of absolute value at most $C + 1$.

Using Lemmas 10 and 11 we now prove Theorem 7.

Proof of Theorem 7. Fix $1 \leq k \leq d-1$. We will bound the number of polynomials $p \in RM(n,d)$ s.t. $wt(p) \leq 2^{-k}(1-\epsilon)$. Let p be any such polynomial. We apply Lemma 10 to p. We can write $(-1)^{p(x)}$ as

$$(-1)^{p(x)} = \mathbb{E}_{a_1,...,a_k \in \mathbb{F}_2^n}[\alpha_{a_1,...,a_k}(-1)^{p_{a_1,...,a_k}(x)}]$$

such that $|\alpha_{a_1,...,a_k}| \leq \frac{10}{\epsilon}$.

We now apply Lemma 11 to the set of polynomials $\{p_{a_1,...,a_k}(x) : a_1, ..., a_k \in \mathbb{F}_2^n\}$ with $\delta = 2^{-(d+2)}$. We get that there are $\ell = O(\frac{d}{\epsilon^2})$ derivatives $\{p_{a_1^i,...,a_k^i} : i \in [\ell]\}$ s.t. the distance between $p(x)$ and $F(x)$ is at most δ, where

$$F(x) = sign(\frac{\sum_{i=1}^{\ell} s_i(-1)^{p_{a_1^i,...,a_k^i}(x)}}{\ell})$$

and $s_1, ..., s_\ell$ are integers of absolute value at most $O(\frac{1}{\epsilon})$.

We now make an important yet simple observation, that will let us bound the number of low weight polynomials by bounding the number of functions $F(x)$. Given any $F(x)$, there can be at most one $p \in RM(n,d)$ s.t. $dist(F,p) \leq \delta$. Assume otherwise that there are two polynomials $p', p'' \in RM(n,d)$ s.t. $dist(p', F) \leq \delta$ and $dist(p'', F) \leq \delta$. By the triangle inequality $dist(p', p'') \leq 2\delta < 2^{-d}$, but this cannot hold if p', p'' are two different polynomials, since the minimum relative distance of $RM(n,d)$ is 2^{-d}.

So, if we bound the number of different functions $F(x)$ of the above form, we will also bound the number of polynomials p of relative weight at most $2^{-k}(1-\epsilon)$. Consider the terms appearing in F:

- We need $\ell = O(\frac{d}{\epsilon^2})$ derivatives and coefficients to describe F completely.
- Any derivative $p_{a_1^i,...,a_k^i}(x)$ is a a polynomial of degree at most $d-k$, and so has at most $2^{\binom{n}{\leq d-k}}$ possibilities.
- Any coefficient s_i has $O(\frac{1}{\epsilon})$ possibilities.

Thus, the total the number of different F's is at most

$$\left(2^{\binom{n}{\leq d-k}} \cdot (1/\epsilon)\right)^{O(\frac{d}{\epsilon^2})} \leq c_1 2^{c_2 \frac{n^{d-k}}{\epsilon^2}}$$

where $c_1 = (1/\epsilon)^{O(d/\epsilon^2)}$ and $c_2 = O(d/(d-k)!)$.

□

We now turn to prove the Lemmas required for the proof of Theorem 7. We prove Lemma 10 in Subsection 2.1 and Lemma 11 in Subsection 2.2.

2.1 Proof of the main technical lemma: Lemma 10

Before proving Lemma 10, we need some claims regarding derivatives. The first claim shows that if a function has non-zero bias, it can be computed by an average of its derivatives.

Claim 12. *Let $g : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function s.t. $bias(g) \neq 0$. Then:*

$$(-1)^{g(x)} = \frac{1}{bias(g)} \mathbb{E}_{a \in \mathbb{F}_2^n}[(-1)^{g_a(x)}]$$

where the identity holds for any $x \in \mathbb{F}_2^n$.

Proof. Fix x. We have:

$$(-1)^{g(x)} \mathbb{E}_{a \in \mathbb{F}_2^n}[(-1)^{g_a(x)}] = \mathbb{E}_{a \in \mathbb{F}_2^n}[(-1)^{g(x)-g_a(x)}] = \mathbb{E}_{a \in \mathbb{F}_2^n}[(-1)^{g(x+a)}] = bias(g)$$

□

The following claim shows that if a function has low weight, then derivatives of it will also have low weight, and thus large bias.

Claim 13. *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function s.t. $wt(f) < 2^{-k}(1-\epsilon)$. Let $a_1, ..., a_s \in \mathbb{F}_2^n$ for $1 \leq s \leq k-1$ be any derivatives, and consider $bias(f_{a_1,...,a_s})$. Then $bias(f_{a_1,...,a_s}) \geq 1 - 2^{s+1-k}(1-\epsilon)$. In particular:*

1. *If $s < k-1$ then $bias(f_{a_1,...,a_s}) \geq 1 - 2^{s+1-k}$*
2. *If $s = k-1$ then $bias(f_{a_1,...,a_s}) \geq \epsilon$*

Proof. Consider $f_{a_1,...,a_s}$

$$f_{a_1,...,a_s} = \sum_{I \subseteq [s]} f(x + \sum_{i \in I} a_i)$$

For random x, the probability that $f(x + \sum_{i \in I} a_i) = 1$ is $wt(f)$, which is at most $2^{-k}(1-\epsilon)$. Thus by union bound,

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[\exists I \subseteq [s],\ f(x + \sum_{i \in I} a_i) = 1] \leq 2^{s-k}(1-\epsilon)$$

In particular it implies that

$$wt(f_{a_1,...,a_s}) = \mathbb{P}_{x \in \mathbb{F}_2^n}[f_{a_1,...,a_s}(x) = 1] \leq 2^{s-k}(1-\epsilon)$$

and we get the bound since $bias(f_{a_1,...,a_s}) = 1 - 2wt(f_{a_1,...,a_s})$. □

We now can prove Lemma 10 using Claims 12 and 13.

Proof of Lemma 10. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a function s.t. $wt(f) \leq 2^{-k}(1-\epsilon)$. Thus $bias(f) = 1 - 2wt(f) > 0$ and by Claim 12 we can write:

$$(-1)^{f(x)} = \frac{1}{bias(f)} \mathbb{E}_{a_1 \in \mathbb{F}_2^n}[(-1)^{f_{a_1}(x)}]$$

If $k = 1$ we are done. Otherwise by Claim 13, f_{a_1} also has positive bias,

$$bias(f_{a_1}) \geq 1 - 2^{s+1-k}(1-\epsilon) > 0$$

and so again by Claim 12 we can write

$$(-1)^{f_{a_1}(x)} = \frac{1}{bias(f_{a_1})} \mathbb{E}_{a_2 \in \mathbb{F}_2^n}[(-1)^{f_{a_1,a_2}(x)}]$$

Thus we have:

$$(-1)^{f(x)} = \frac{1}{bias(f)} \mathbb{E}_{a_1 \in \mathbb{F}_2^n}[\frac{1}{bias(f_{a_1})} \mathbb{E}_{a_2 \in \mathbb{F}_2^n}[(-1)^{f_{a_1,a_2}(x)}]]$$

We can continue this process as long as we can guarantee that $f_{a_1,...,a_s}$ has non-zero bias for all $a_1, ..., a_s \in \mathbb{F}_2^n$. By Claim 13 we know this happens for $s \leq k-1$, and thus we have:

$$(-1)^{f(x)} = \mathbb{E}_{a_1,...,a_k \in \mathbb{F}_2^n}[\alpha_{a_1,...,a_k}(-1)^{f_{a_1,...,a_k}(x)}]$$

where

$$\alpha_{a_1,...,a_k} = \frac{1}{bias(f)} \frac{1}{bias(f_{a_1})} \frac{1}{bias(f_{a_1,a_2})} \cdots \frac{1}{bias(f_{a_1,...,a_{k-1}})}$$

We now bound $\alpha_{a_1,...,a_k}$. By Claim 13 we get that:

$$\alpha_{a_1,...,a_k} \leq \frac{1}{\epsilon} \prod_{s=1}^{k-2} \frac{1}{1-2^{s-k+1}} \leq \frac{1}{\epsilon} \prod_{r \geq 1} \frac{1}{1-2^{-r}} \leq \frac{10}{\epsilon}$$

□

2.2 Proof of Approximation by sampling Lemma: Lemma 11

Proof of Lemma 11. Choose $h_1, ..., h_\ell$ uniformly and independently from H. Fix $x \in \mathbb{F}_2^n$, and let Z_i be the random variable

$$Z_i = c_{h_i}(-1)^{h_i(x)}$$

and let $S = \frac{Z_1 + ... + Z_\ell}{\ell}$. We will use the fact that if $|S - (-1)^{f(x)}| < 1$ then $sign(S) = (-1)^{f(x)}$. We first bound the probability that

$$|S - (-1)^{f(x)}| > 1/4$$

By regular Chernoff arguments for bounded independent variables, since $\mathbb{E}[S] = (-1)^{f(x)}$ and each Z_i is of absolute value of at most C, we get that

$$\mathbb{P}_{h_1,...,h_\ell \in H}[|S - (-1)^{f(x)}| > 1/4] \leq e^{-\frac{\ell}{32C^2}}$$

(see for example Theorem A.1.16 in [1]).

In particular for $\ell = O(C^2 \log 1/\delta)$ we get that

$$\mathbb{P}_{h_1,...,h_\ell \in H}[|S - (-1)^{f(x)}| > 1/4] \leq \delta$$

Thus by averaging arguments, there exists $h_1, ..., h_\ell$ s.t.

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[|\frac{c_{h_1}(-1)^{h_1(x)} + ... + c_{h_\ell}(-1)^{h_\ell(x)}}{\ell} - (-1)^{f(x)}| \geq 1/4] \leq \delta$$

We now round each coefficient to a close rational, without damaging the approximation error. The coefficient of $(-1)^{h_i(x)}$ is $\alpha_i = \frac{c_{h_i}}{\ell}$. If we round c_{h_i} to the closest integer $[c_{h_i}]$, we get that the coefficient of each $(-1)^{h_i(x)}$ is changed by at most $\frac{1}{2\ell}$, and thus the total approximation is changed by at most $1/2$. Hence we have:

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[|\frac{[c_{h_1}](-1)^{h_1(x)} + ... + [c_{h_\ell}](-1)^{h_\ell(x)}}{\ell}) - (-1)^{f(x)}| \geq 3/4] \leq \delta$$

Thus we got that

$$\mathbb{P}_{x \in \mathbb{F}_2^n}[sign(\frac{[c_{h_1}](-1)^{h_1(x)} + ... + [c_{h_\ell}](-1)^{h_\ell(x)}}{\ell}) \neq (-1)^{f(x)}] \leq \delta$$

□

3 List-decoding size of Reed-Muller codes

In this section we turn to the problem of bounding the list-decoding size of Reed-Muller codes, and we prove the First Main Theorem (Theorem 4). We will see that the same techniques we used in Section 2 to bound the weight distribution, can be applied with minor variants to also bound the list-decoding size.

The list-decoding size of a code is at least the accumulative weight distribution, i.e. $L(\alpha) \geq A(\alpha)$. However, the list-decoding size can sometimes be much larger than the accumulative weight distribution.

Theorem 4 is a direct corollary of Theorem 14, giving an upper bound on the list-decoding size at distance $2^{\ell-d} - \epsilon$, and the same lower bound we used to bound the accumulative weight distribution, obtained in Lemma 8.

Theorem 14 (Upper bound on the list-decoding size)**.** *For any integer $1 \leq k \leq d-1$,*

$$L(2^{-k}(1-\epsilon)) \leq c_1 2^{c_2 \frac{n^{d-k}}{\epsilon^2} + c_3 \frac{n}{\epsilon^2}}$$

where $c_1 = (1/\epsilon)^{O(d/\epsilon^2)}$, $c_2 = O(d/(d-k)!)$ and $c_3 = O(dk)$. Importantly, c_1, c_2, c_3 are independent of n, and c_2, c_3 are independent of ϵ. In particular for constant d we get that

$$L(2^{-k} - \epsilon) \leq 2^{O(\frac{n^{d-k}}{\epsilon^2})}$$

Proof of Theorem 14. The proof will be similar to the proof of Theorem 7. Fix $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to be any function. We will bound the number of polynomials p of degree at most d s.t. $dist(p, f) \leq 2^{-k}(1-\epsilon)$. Let $p \in RM(n,d)$ be such a polynomial, i.e. $dist(p,f) \leq 2^{-k}(1-\epsilon)$. Let $g(x) = p(x) - f(x)$, then $wt(g) \leq 2^{-k}(1-\epsilon)$. As in the proof of Theorem 7, we use the derivatives of g to approximate g. Set $\delta = 2^{-(d+2)}$. By Lemma 10 there are $\ell = O(\frac{d}{\epsilon^2})$ derivatives $\{g_{a_1^i,\ldots,a_k^i} : i \in [\ell]\}$ s.t. the distance between $g(x)$ and $F(x)$ is at most δ, where

$$F(x) = sign(\frac{\sum_{i=1}^{\ell} s_i (-1)^{g_{a_1^i,\ldots,a_k^i}(x)}}{\ell})$$

Thus we have that $F + f$ approximates p, since:

$$dist(p, F+f) = dist(p-f, F) \leq \delta$$

As in the proof of Theorem 7, given F (and f) there can be at most a single $p \in RM(n,d)$ s.t. $dist(p, F+f) \leq \delta$, and so if we will bound the number of functions F we will bound the number of codewords close to f.

Consider the derivative $g_{a_1^i,\ldots,a_k^i}(x)$ used in the expression for F. By linearity of derivation it can be decomposed as

$$g_{a_1^i,\ldots,a_k^i}(x) = p_{a_1^i,\ldots,a_k^i}(x) - f_{a_1^i,\ldots,a_k^i}(x)$$

Each $p_{a_1^i,\ldots,a_k^i}(x)$ is a degree $d-k$ polynomial, and so has at most $2^{\binom{n}{\leq d-k}}$ possibilities. Each $f_{a_1^i,\ldots,a_k^i}(x) = \sum_{S \subseteq [k]} f(x + \sum_{j \in S} a_j^i)$ can be described by the values of $a_1^i, \ldots, a_k^i \in \mathbb{F}_2^n$, since we have access to f, and so has at most 2^{kn} possibilities. Each coefficient s_i has $O(1/\epsilon)$ possibilities. Thus, in total the number of different F's is at most

$$\left(2^{\binom{n}{\leq d-k}+kn} \cdot (1/\epsilon)\right)^{O(\frac{d}{\epsilon^2})} \leq c_1 2^{c_2 \frac{n^{d-k}}{\epsilon^2} + c_3 \frac{n}{\epsilon^2}}$$

where $c_1 = (1/\epsilon)^{O(d/\epsilon^2)}$, $c_2 = O(d/(d-k)!)$ and $c_3 = O(kd)$. □

Acknowledgement. The second author would like to thank his advisor, Omer Reingold, for ongoing advice and encouragement. He would also like to thank Microsoft Research for their support during his internship.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, Second edition, published by John Wiley, 2000.

[2] A. Bogdanov and E. Viola. Pseudorandom bits for polynomials via the Gowers norm. In *the 48th Annual Symposium on Foundations of Computer Science (FOCS 2007).*

[3] O. Goldreich and L. Levin, *A hard core predicate for all one way functions*, In the Proceedings of the 21st ACM Symposium on Theory of Computing (STOC), 1989.

[4] P. Gopalan, A. Klivans and D. Zuckerman, *List-Decoding Reed Muller Codes over Small Fields*, In the Proceedings of the 40th ACM Symposium on Theory of Computing (STOC), 2008.

[5] V. Guruswami, *List decoding of Error-Correcting Codes*, vol 3282 of Lecture notes in Computer Science, Springer 2004.

[6] T. Kaufman and S. Lovett, *Worst case to Average Case Reductions for Polynomials*, To appear in the Proceedings of the 49th Annual Symposium on Foundations of Computer Science (FOCS), 2008.

[7] E. Kushilevitz and Y. Mansour, *Learning Decision Trees using the Fourier Spectrum*, SIAM Journal of Computing, 22(6), (1993), pp 1331-1348.

[8] T. Kasami and N. Tokura, *On the weight structure of Reed-Muller codes*, In the IEEE Transactions on Information Theory 16 (Issue 6), 1970.

[9] J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, Amsterdam, North-Holland, 1977.

[10] M. Sudan, *Decoding of Reed-Solomon codes beyond the error-correction bound*, Journal of Complexity, 13, (1997), pp. 180-193.

[11] M. Sudan, *List decoding: Algorithms and Applications* , SIGACT News, 31 (2000), pp 16-27.

[12] M. Sudan, L. Trevian, S. Vadhan *Pseudorandom Generators without the XOR Lemma*, J. Comput. Syst. Sci., 61 (2001), pp 236-266.

A Generalized Reed-Muller codes

The problems of bounding both the accumulative weight distribution and the list-decoding size can be extended to Generalized Reed-Muller, the code of low-degree polynomials over larger fields. However, our techniques fail to prove tight result in these cases. We briefly describe the reasons below, and give some partial results.

We start by making some basic definitions. Let q be a prime, and let $GRM_q(n,d)$ denote the code of multivariate polynomials $p(x_1,...,x_n)$ over the field $\mathbb{F}_q$, of total degree at most d.

Definition 7. The relative weight of a function $f:\mathbb{F}_q^n \to \mathbb{F}_q$ is the fraction of non-zero elements,

$$wt(f) = \frac{1}{q^n}|\{x \in \mathbb{F}_q^n : f(x) \neq 0\}|$$

Definition 8. The relative distance between two functions $f,g:\mathbb{F}_q^n \to \mathbb{F}_q$ is defined as

$$dist(f,g) = \mathbb{P}_{x\in\mathbb{F}_q^n}[f(x) \neq g(x)]$$

The accumulative weight distribution and the list-decoding size are defined analogously for $GRM_q(n,d)$, using the appropriate definitions for relative weight and relative distance. We denote them by A_q and L_q. For each $1 \leq k \leq d$, we define a distance r_k:

1. For $k=1$, let $d=(q-1)a+b$, where $1 \leq b \leq q-1$. Define $r_1 = q^{-a}(1-b/q)$.
2. For $2 \leq k \leq d-1$, let $d-k=(q-1)a+b$, where $1 \leq b \leq q-1$. Define $r_k = q^{-a}(1-b/q)(1-1/q)$.
3. For $k=d$, define $r_d = 1-1/q$.

We conjecture that both for the accumulative weight distribution and the list-decoding size, the distances r_k are the thresholds for the exponential dependency in n:

Conjecture 15. *Let $\epsilon > 0$ be constant, and consider $GRM_q(n,d)$ for constant d. Then:*

- *For $\alpha \leq r_1 - \epsilon$ both $A_q(\alpha)$ and $L_q(\alpha)$ are constants.*
- *For $r_k \leq \alpha \leq r_{k+1} - \epsilon$ both $A_q(\alpha)$ and $L_q(\alpha)$ are $2^{\Theta(n^k)}$.*
- *For $\alpha \geq r_d$ both $A_q(\alpha)$ and $L_q(\alpha)$ are $2^{\Theta(n^d)}$.*

Proving lower bounds for $A_q(r_k)$ is similar to the case of $RM(n,d)$.

Lemma 16 (Lower bound for A_q)**.** *For any integer $1 \leq k \leq d$,*

$$A_q(r_k) \geq 2^{\Omega(n^k)}$$

The problem is proving matching upper bounds. Using directly the derivatives method we used to give upper bounds for $RM(n,d)$ gives the same bounds for $GRM_q(n,d)$, alas they are not tight for $q > 2$:

$$A_q(2^{-k} - \epsilon) \leq 2^{O(n^{d-k})}$$

If we would like to get upper bounds closer to the lower bounds, a natural approach would be to generalize Lemma 10 to taking several derivatives in the same direction (which is possible over larger fields). This would give us tight results for some values of k, if we could also generalize Claim 12 to the case of taking a multiple derivative in the same direction. However, we didn't find a way of doing so.

Instead, we give partial results for Conjecture 15 in the two ends of the scale: when $\alpha \leq r_1 - \epsilon$, and when $r_{d-1} \leq \alpha \leq r_d - \epsilon$ (when $\alpha \geq r_d$ Lemma 16 gives $L_q(\alpha)$ and $A_q(\alpha)$ are both exponential in n^d).

First, the minimal distance of $GRM_q(n,d)$ is known to be r_1. Thus, for any $\epsilon > 0$, $A_q(r_1 - \epsilon) = 1$. Gopalan, Klivans and Zuckerman [4] prove that $L_q(r_1 - \epsilon)$ is constant when $q-1$ divides d:

Theorem 17 (Corollary 18 in [4])**.** *Assume $q-1$ divides d. Then:*

$$L_q(r_1 - \epsilon) \leq c(q,d,\epsilon)$$

Moving to the case of $r_{d-1} \leq \alpha \leq r_d - \epsilon$, we prove:

Lemma 18. *Let $\epsilon > 0$ be constant. then:*

$$A_q(r_d - \epsilon) \leq 2^{O(n^{d-1})}$$

We now move on to prove Lemmas 16 and 18. We start with Lemma 16:

Proof of Lemma 16. We start by proving for $2 \leq k \leq d-1$. Let $d - k = (q-1)a + b$, where $1 \leq b \leq q-1$. Single out $a+2$ variables $x_1, ..., x_{a+2}$, and let g be any degree k polynomial on the remaining variables. The following polynomial has degree d and weight exactly $q^{-a}(1-b/q)(1-1/q)$:

$$g'(x_1, ..., x_n) = \left(\prod_{i=1}^{a} \prod_{j=1}^{q-1} (x_i - j) \right) \left(\prod_{j=1}^{b} (x_{a+1} - j) \right) (x_{a+2} + g(x_{a+3}, ..., x_n))$$

The number of distinct polynomial g is $2^{\Omega(n^d)}$.

The proofs for $k = 1$ and $k = d$ are similar: for $k = 1$, let $d = (q-1)a + b$. Let $l_1(x), ..., l_{a+1}(x)$ be any independent linear functions, and consider

$$g'(x_1, ..., x_n) = \left(\prod_{i=1}^{a}\prod_{j=1}^{q-1}(l_i(x) - j)\right)\left(\prod_{j=1}^{b}(l_{a+1}(x) - j)\right)$$

For $k = d$, let g be any degree d polynomial on variables $x_2, ..., x_n$, and consider $g'(x_1, ..., x_n) = x_1 + g(x_2, ..., x_n)$. □

We now continue to prove Lemma 18. We first make some necessary definitions.

Definition 9. The bias of a polynomial $p(x_1, ..., x_n)$ over $\mathbb{F}_q$ is defined to be

$$bias(p) = \mathbb{E}_{x \in \mathbb{F}_q^n}[\omega^{p}(x)]$$

where $\omega = e^{2\pi i/q}$ is a primitive q-th root of unity.

Kaufman and Lovett [6] prove that biased low-degree polynomials can be decomposed into a function of a constant number of lower degree polynomials:

Theorem 19 (Theorem 2 in [6])**.** *Let $p(x_1, ..., x_n)$ be a degree d polynomial, s.t. $|bias(p)| \geq \epsilon$. Then p can be decomposed as a function of a constant number of lower degree polynomials:*

$$p(x) = F(g_1(x), ..., g_c(x))$$

where $deg(g_i) \leq d - 1$, and $c = c(q, d, \epsilon)$.

We will use Theorem 19 to bound $A(r_d - \epsilon)$ for any constant $\epsilon > 0$.

Proof of Lemma 18. We will show that any polynomial $p \in GRM_q(n, d)$ s.t. $wt(p) \leq 1 - 1/p - \epsilon$ can be decomposed as

$$p(x) = F(g_1(x), ..., g_c(x))$$

where $deg(g_i) \leq d - 1$, and c depends only on q, d and ϵ. Thus the number of such polynomials is bounded by the number of possibilities to choose c degree $d - 1$ polynomials, and a function $F : \mathbb{F}_q^c \to \mathbb{F}_q$. The number of such possibilities is at most $2^{O(n^{d-1})}$. Let p be s.t. $wt(p) \leq 1 - 1/p - \epsilon$. We will show there exists $\alpha \in \mathbb{F}_q$, $\alpha \neq 0$ s.t. $bias(\alpha p) \geq \epsilon$. We will then finish by using Theorem 19 on the polynomial αp.

Consider the bias of αp for random $\alpha \in \mathbb{F}_q$:

$$\mathbb{E}_{\alpha \in \mathbb{F}_q}[bias(\alpha p)] = \mathbb{E}_{\alpha \in F_q, x \in \mathbb{F}_q^n}[\omega^{\alpha p(x)}] = 1 - wt(p)$$

since for x's for which $p(x) = 0$, $\mathbb{E}_{\alpha \in F_q}[\omega^{\alpha p(x)}] = 1$, and for x s.t. $p(x) \neq 0$, $\mathbb{E}_{\alpha \in F_q}[\omega^{\alpha p(x)}] = 0$. We thus get that:

$$\mathbb{E}_{\alpha \in \mathbb{F}_q \setminus \{0\}}[bias(\alpha p)] = 1 - \frac{q}{q-1}wt(p) \geq \frac{q}{q-1}\epsilon$$

So, there must exist $\alpha \neq 0$ s.t. $bias(\alpha p) \geq \epsilon$. □